Monitoring Variability in Complex Manufacturing Process: Data Analysis Viewpoint with Application

Maman Abdurachman Djauhari[1], Norhaidah Mohd Asrah[2], Irianto[3], Noor Akma Ibrahim[4]

[1]*Graduate School, Indonesian Institute of Education (IPI)*
Jl. Terusan Pahlawan 32, Garut 44151, Indonesia
[1] maman_djauhari@yahoo.com
[2]*Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia*
Pagoh Campus, 84600 Pagoh, Johor, Malaysia
[2] norhaida@uthm.edu.my
[3&4]*Institute for Mathematical Research, Universiti Putra Malaysia*
43400 UPM Serdang, Selangor, Malaysia
[3] GS43315@student.upm.edu.my
[4] nakma@upm.edu.my

Abstract— **To relate the control limits of Shewhart-type chart to the p-value, the control charting techniques were constructed based on statistical inference scheme. However, in daily practice of complex process variability (CPV) monitoring operation, these limits have nothing to do with the p-value. We cannot put any number to p. Instead, p is just read as "most probably". These words mean that in practice we are finally working under data analysis scheme instead. For this reason, in this paper we introduce the application of STATIS in CPV monitoring operation. It is a data analysis method to label the sample(s) where anomalous covariance structure occurs. This method is algebraic in nature and dominated by principal component analysis (PCA) principles. The relative position of a covariance matrix among others is visually presented along the first two eigenvalues of the so-called "scalar product matrix among covariance matrices". Its strength will be illustrated by using a real industrial example and the results, compared with those given by the current methods, are very promising. Additionally, root causes analysis is also provided. However, since STATIS is a PCA-like, it does not provide any control chart, i.e., the history of process performance. It is to label the anomalous sample(s). To the knowledge of the authors, the application of STATIS in complex statistical process control is an unprecedented. Thus, it will enrich the literature of this field.**

Keywords— *conjoint analysis, Escoufier's operator, generalized variance, Hilbert-Schmidt space, vector variance.*

International Journal of Supply Chain Management
IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)
Copyright © ExcelingTech Pub, UK (http://excelingtech.co.uk/)

1. Introduction

The word "complex" in CPV refers to several interrelated critical to qualities (CTQs). It is customarily replaced by the word "multivariate" when the interrelationship among CTQs is quantified in terms of Pearson correlation. It represents the mathematical and statistical complexity faced in monitoring CPV and generally in complex statistical process control (CSPC). As mentioned in [19] and repeated in [20], [21], monitoring CPV is as important as monitoring process target.

Although it is an important part in CSPC, we learn that the technique for monitoring CPV is still in development. This is perhaps due to the fact that (1) it requires large sample size, and (2) CPV is difficult to measure. And among the very limited number of CPV measures available, according to [25], generalized variance (say GV for brevity) is the most adopted measure. We can find its importance as a CPV measure in, among others, [1], [12], [18], [23], [27], [28], [29], and [30].

However, as mentioned in [5], it is unfortunate that GV converges in distribution very slow to normality. It is then not favorable to be used in manufacturing industrial practice where only small sample is available. And when dealing with small sample, in the current practice, the control limits do not reflect the *p*-value. In other words, they have nothing to do with probability of false alarm (PFA). As a consequence, *p* cannot be represented by any number or percentage and in practice we are happy to read p as "most probably". See [20]-[21] for the details.

To have a better method for CPV monitoring, [7] introduced a technique constructed based on vector variance (VV). The notion of VV is originally appeared in [8] and introduced as CPV measure by [4]. The purpose of VV-based method is to reduce the weakness of GV-based method. See also [18] and [26] for a discussion of these methods. Since VV converges in distribution faster than GV to normality, see [6], this method needs less sample size. In addition, VV-based method is more sensitive to the small shift in covariance structure. However, see [5], it is still not favorable in manufacturing industrial practice because, for small sample, the control limits have still nothing to do with the p-value; they do not reflect PFA.

These two methods are constructed based on the theory of statistical inference in order to have direct relationship between the control limits and the p-value. But finally, p must be interpreted as "most probably". This means that, in practice, these methods are not an inferential method but a data analysis or also called data exploration method. This is the point that motivates us to write this paper.

This point of view leads us to introduce here the application of what French statisticians called STATIS in CSPC to enrich data exploration results. STATIS stands for "Structuration des Tableaux à Trois Indices de la Statistique". It was first introduced in a thesis by [16] in order to do conjoint analysis of three-way data matrix. Since then, its development and application in many areas can be found in, for example, [10], [15] and [17]. This method is algebraic in nature and constructed based on PCA principles. For those who are interested, the suggested reading in French is [9], and [13], and in English is [11] and [15].

We show that, through graphical representation, STATIS could label the sample(s) where anomalous covariance structure is present. In addition, it could lead to identify the root causes why at a given sample the covariance structure has been shifted. This facility cannot be provided by the two Shewhart-type methods mentioned above. To illustrate its strength, a real industrial example will be presented, and the results will be compared with those given by GV-based and VV-based methods.

In the rest of the paper our discussion will be organized as follows. We start in the next section by recalling GV-based control chart (GV-chart for brevity); its theoretical background and practical implementation will be highlighted. Later on, in Section 3 we show the performance of convergence in distribution of GV and that of VV to normality. Due to slow convergence, in Section 4 we adopt the method of STATIS in CPV monitoring. Here, our focus is on its implementation without going into the details of theoretical background. In Section 5 a real example coming from the Centre for Indonesian Army Industry, Ltd., will be presented and discussed. Finally, closing remarks in the last section will end this work.

2. The Most Adopted Shewhart-Type Chart

Monitoring operation of CPV is basically conducted based on m independent samples of the same size n. Statistically speaking, it is equivalent to testing repeatedly the following $H_0 : \Sigma_i = \Sigma_0$ for all $i = 1,2,\ldots,m$, versus $H_1 : \Sigma_i \neq \Sigma_0$ for an i in the set $\{1,2,\ldots,m\}$. Here, Σ_i is the covariance matrix of the population where the i-th sample is drawn and Σ_0 is the hypothetical covariance matrix. The p-value, also called PFA, is set to be 0.0027 or 0.27% while the two sided critical values represent the lower control limit (LCL) and upper control limits (UCL).

If S_i is the i-th sample covariance matrix, under $H_0 : \Sigma_i = \Sigma_0$ for all $i = 1,2,\ldots,m$, [2], see also [22], shows that,

$$\sqrt{n}\left(|S_i| - |\Sigma_0|\right) \xrightarrow{d} N\left(0, 2p|\Sigma_0|^2\right) \qquad (2.1)$$

We show that the parameters $|\Sigma_0|$ and $\frac{2p}{n}|\Sigma_0|^2$ in (2.1) are not the true mean and true variance of $|S|$. They are the limit of the true ones when n tends to infinity. The true mean and variance are,

$$E(|S_i|) = b_1|\Sigma_0| \quad \text{and} \quad Var(|S_i|) = b_2|\Sigma_0|^2,$$

where,

$$b_1 = \frac{1}{(n-1)^p} \prod_{k-1}^{p} (n-k)$$

and

$$b_2 = b_1\left\{\frac{1}{(n-1)^p}\prod_{k=1}^{p}(n-k+2)-b_1\right\}.$$

That is the reason why (2.1) is not used in the literature to construct the GV-chart. Instead, as can be seen in the next sub-section, (2.2) is used

$$\frac{1}{\sqrt{b_2}}\left(|S_i|-b_1|\Sigma_0|\right)\xrightarrow{d} N\left(0,|\Sigma_0|^2\right) \qquad (2.2)$$

In addition, the convergence in (2.1) is slower than that in (2.2)

2.1 GV-chart

Suppose m independent samples each of size n drawn from a p-variate normal distribution with positive definite covariance matrix Σ are available for CPV monitoring operation. We consider again S_i the covariance matrix of the i-th sample; $i = 1,2,\ldots,m$. GV-chart is constructed by plotting on the same diagram the sample GV, $|S_i|$; $i = 1,2,\ldots,m$, and the control limits. In this paragraph, an evolution of GV-chart will be presented to clarify its usefulness.

First of all, no literature brings (2.1) into practice in CPV monitoring because the parameters mean, and variance are not the true ones. Instead, literature uses (2.2). In this case, see [20]-[21], the control limits $UCL = \frac{|\bar{S}|}{b_1}\left(b_1 + 3\sqrt{b_2}\right)$, $CL = |\bar{S}|$, and $LCL = \max\left\{0, \frac{|\bar{S}|}{b_1}\left(b_1 - 3\sqrt{b_2}\right)\right\}$ are usually used. Here, $\bar{S}$ is the average of S_i. It is worth noting that, in practice, these control limits are used regardless whether the sample size *n* is large or small.

Of course, these control limits are better than those (if available) given by (2.1) since, as mentioned above, the convergence of (2.1) is slower than that of (2.2). However, see [3], they are not unbiased. This author shows that when we deal with *m* independent samples, under H_0 we have

$$\frac{1}{\sqrt{b_4}}\left(|\bar{S}|-b_3|\Sigma_0|\right)\xrightarrow{d} N\left(0,|\Sigma_0|^2\right) \qquad (2.3)$$

where, $b_3 = \frac{1}{(m(n-1))^p}\prod_{k=1}^{p} m(n-1)-k+1$ and

$$b_4 = b_3\left\{\frac{1}{(m(n-1))^p}\prod_{j=1}^{p}[m(n-1)-j+3]-b_3\right\}.$$

Accordingly, (2.3) provides these unbiased control limits $UCL = |\bar{S}|\left(\frac{b_1}{b_3}+3\sqrt{\frac{b_2}{b_3^2+b_4}}\right)$, $CL = |\bar{S}|\frac{b_1}{b_3}$, and $LCL = \max\left\{0, |\bar{S}|\left(\frac{b_1}{b_3}-3\sqrt{\frac{b_2}{b_3^2+b_4}}\right)\right\}$. This is the latest form of GV-chart evolution.

That is the scenario at a glance of GV-chart evolution. It is worth noting that whichever the control limits we use, if *n* is sufficiently large, the *p*-value is 0.27%. Meanwhile, for small *n*, *p* must be read as "most probably" and does not represent the PFA.

We conclude that no matter whether the sample size is large or small, the control chart is always the same. The only difference lies in the way we interpret the chart. When *n* is sufficiently large, the chart shows graphically how to make decision in testing the hypothesis H_0 versus H_1. On the other hand, when *n* is small, this chart has nothing to do with hypothesis testing; it must be considered as a data analysis or data exploration tool.

2.2 VV-chart

Under H_0, [6] show that for each sample i,

$$\frac{n-1}{\sqrt{8n}}\left\{Tr\left(S_i^2\right)-\frac{n+1}{n-1}Tr\left(\Sigma_0^2\right)\right\}\xrightarrow{d} N\left(0,Tr\left(\Sigma_0^4\right)\right) \qquad (2.4)$$

Here, Tr is the trace operator on a square matrix, i.e., the sum of all diagonal elements. The VV-chart is then constructed by plotting on the same diagram the value of $Tr\left(S_i^2\right)$ and that of control limits. [6] show that (2.4) leads to $UCL = \frac{n+1}{n-1}Tr\left(\bar{S}^2\right)+3\sqrt{\frac{8n}{(n-1)^2}Tr\left(\bar{S}^4\right)}$, $CL = \frac{n+1}{n-1}Tr\left(\bar{S}^2\right)$ and $LCL = \max\left\{0, \frac{n+1}{n-1}Tr\left(\bar{S}^2\right)-3\sqrt{\frac{8n}{(n-1)^2}Tr\left(\bar{S}^4\right)}\right\}$ regardless the sample size n. Interestingly, they are unbiased.

Like GV-chart, when *n* is sufficiently large, VV-chart is another graphical representation on how to make decision in testing the hypothesis H_0 versus

H_1 with p-value 0.27%. And, when n is small, it is a data analysis tool and cannot be used to do hypothesis testing.

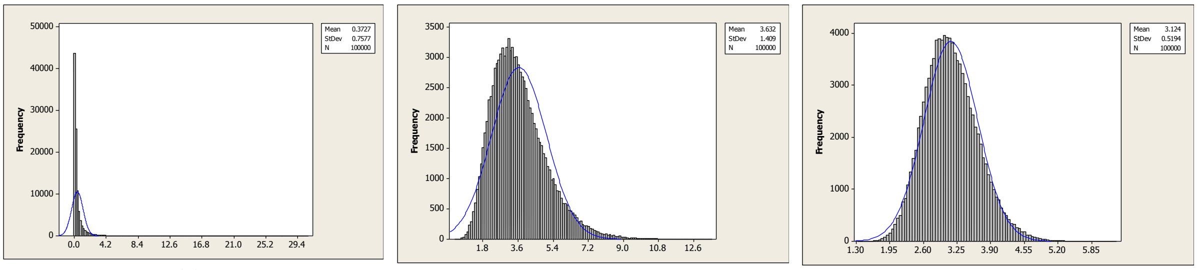
3. Convergence Performance

To illustrate the convergence performance of GV and that of VV, simulation experiments were conducted for selected value of n and p (number of CTQs). In what follows we report the results for $p = 3$ and $n = 5$, 20 and 100 (representing small, moderate and large sample size). In Figure 1 and Figure 2 the histogram of GV and that of VV issued from 100,000 averages of simulated data are presented, respectively. Figure 1(a) is for $n = 5$ while Figures 1(b) and 1(c) are for $n = 20$ and 100.

This figure illustrates that, for $n = 5$, the empirical distribution of GV is strongly skewed to the right and far from normality. According to Anderson-Darling's test, see [24], $AD = 14561.56$ and p-value < 0.005. It is also so for moderate $n = 20$ ($AD = 2508.491$ and p-value < 0.005). Even for large $n = 100$, see Figure 1(c), the distribution is still far from being normal ($AD = 498.787$ and p-value < 0.005).

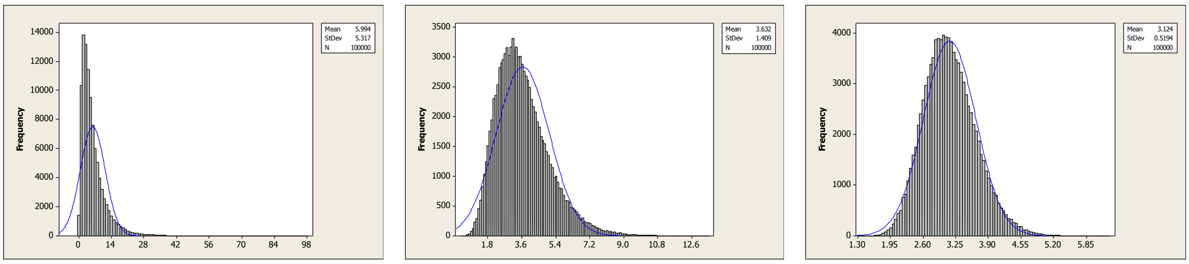
Figure 2(a) is the histogram of VV for $n = 5$ while Figures 2(b) and 2(c) are for $n = 20$ and $n = 100$. Here also, similar situation revealed. For large value of n = 100, the histogram is seemingly close to normality. However, AD = 485.514 and p-value < 0.005 indicate that it is still far from normality but better than GV which has AD = 498.787.

A more general result showing that GV and VV converge very slowly to normality is given in [7]. This strengthens our claim that in practice both GV-chart and VV-chart must be considered as data analysis tools and not related to hypothesis testing H_0 versus H_1. This motivates us to introduce in the next section the application of a data analysis method called STATIS to enrich data exploration results.



(a) (b) (c)

Figure 1. Histogram of GV for $p = 3$ and $n = 5$ (a), 20 (b) and 100 (c)



(a) (b) (c)

Figure 2. Histogram of VV for $p = 3$ and $n = 5$ (a), 20 (b) and 100 (c)

4. How STATIS Works in CPV Monitoring

This method was first introduced in [16] and developed based on the notion of Escoufier's operator related to a data matrix. Basically, see [8], the set of all such operators completed with scalar product defined by Trace is a Hilbert-Schmidt space; the sum of all eigenvalues of each operator is finite. This allows us to transform the study of a sequence of m independent sample covariance matrices into that of m independent operators. And its computation is made practical since the scalar product of two operators is equal to the trace of the multiplication of related covariance matrices. Under this model, CPV monitoring operation can be considered as labeling process of anomalous covariance matrix. This is what STATIS is for.

Consider again the sequence of m covariance matrices S_i; $i = 1,2,\ldots,m$. Let us write C a matrix of size $(m \times m)$ where its general element is defined by $c_{ij} = Tr(S_i S_j)$. The diagonalization of C gives us a vector representation of the matrices S_i on the first two principal components. It is based on this representation that we analyze the relative position of these matrices. An example in the next section will clarify how it works.

5. Industrial Example

We discuss the industrial example presented in [3]. The data are collected during flange manufacturing process at Centre of Indonesian Army Industry located in Bandung, Indonesia. The number of CTQs is p = 3 and the sample size is n = 5.

5.1 Result Issued from GV-Chart and VV-Chart

Figure 3 presents (a) GV-chart and (b) VV-chart without UCL. Since n is very small, there is no significant role of UCL in making the decision.

From this figure we conclude that, according to both charts, most probably the sample 16 is the strongest suspect. Furthermore, GV-chart shows that sample 6 is potential to be suspect while VV-chart indicates that sample 3 is also another potential suspect. Now, let us see what STATIS can do for us.

5.2 Evidence from STATIS

On the other hand, the results from STATIS are shown in Figures 4 and 5.

These figures strongly indicate that at samples 3, 6, and 16 the covariance structure has been shifted. The shift can be more clearly seen in Figures 4-5 than in Figure 3.

Further analysis shows in Figure 6(a) the run chart of the variance of the first variable (blue), second (orange) and third (grey). Meanwhile, Figure in 6(b) is the run chart of the covariance of the first and the second variables in blue, that of the first and the third in orange, and that of the second and the third in grey.

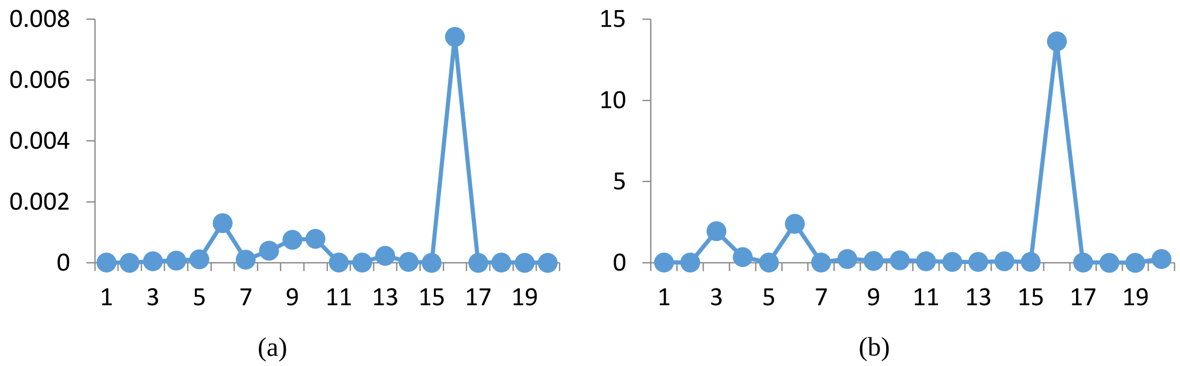


Figure 3. GV-chart (a) and VV-chart (b)

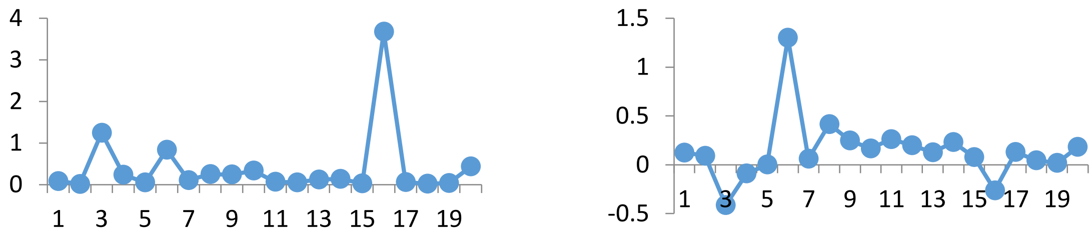
From Figure 6 we see that,

1. At sample 16, covariance structure has been changed due to the change of the variance of the first variable, the covariance of the first and the second variables, and the covariance of the first and the third variables.

2. At sample 6, covariance structure has been changed due to the change of the variance of the third variable, and that of the covariance of the first and the third variables.

3. At sample 3, covariance structure has been changed due to the change of the variance of the first variable.

To close this section, it is worthwhile to note that, as can be seen in Figure 7, the change of covariance structure is not due to the change of correlations.



(a) (b)

FIGURE 4. Run chart along (a) the first and (b) the second principal components

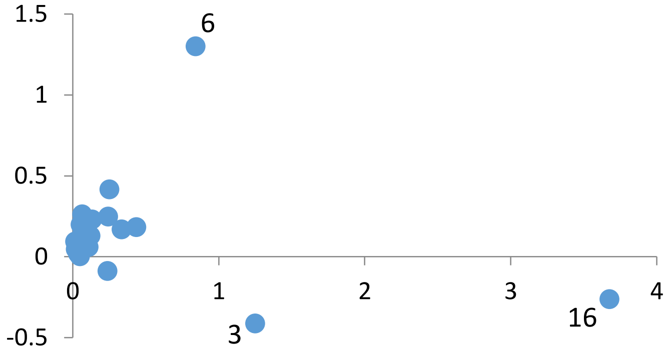
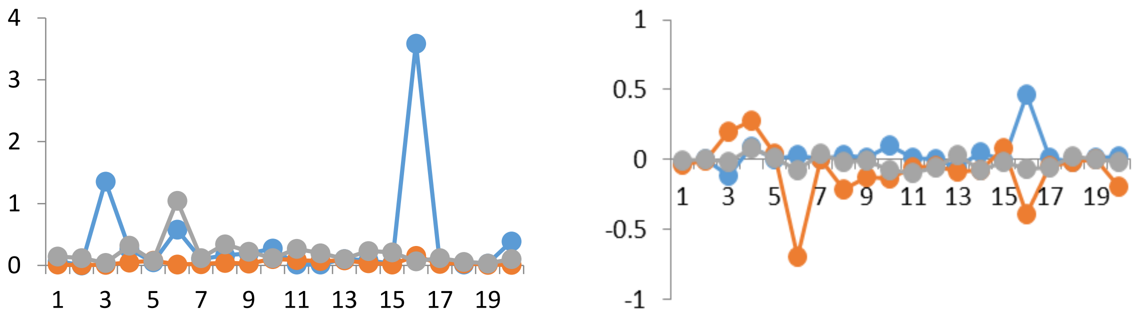


FIGURE 5. Representation of S_i along the first two principal components



(a) (b)

FIGURE 6. Run chart of the three variances (a) and the three covariances (b)

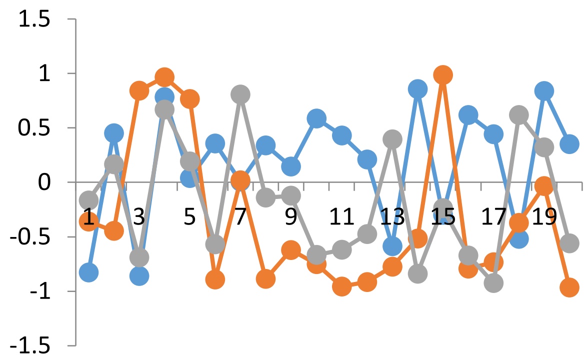


FIGURE 7. Run chart of the three correlations

6. Closing Remarks

We show that, in practice, Shewhart-type charts for CPV monitoring such as GV-chart and VV-chart have nothing to do with testing hypothesis. The control limits have nothing to do with *p*-value and we cannot put any number to *p*. Accordingly, these charts which originally constructed for statistical inference purpose finally become data analysis tool and not related to inferential analysis.

In data analysis scheme, to enrich the results of data exploration, the application of STATIS has been introduced and the results are very promising. An industrial example has successfully illustrated the power of STATIS.

Acknowledgments

The authors are very grateful to the Editor and anonymous referees for their comments that led to this presentation. They acknowledge Universiti Tun Hussein Onn Malaysia for the sponsorship under TIER 1 research grant, 2017, with vote number U902. Special thanks go to Graduate School, Indonesian Institute of Education (IPI), and Centre for Research in Statistics and Data Analysis, Bandung, Indonesia, for providing the first author with research facilities and library access, Faculty of Applied Science and Technology (FAST), Universiti Tun Hussein Onn Malaysia, for providing the second author with all facilities and Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia, for providing the third and fourth authors with research facilities.

References

[1] F.B. Alt and G.E. Bedewi, "*SPC for Dispersion for Multivariate Data*", ASQC Quality Congress Transactions. Milwaukee, WI: American Society for Quality Control, 1986.

[2] T.W. Anderson, *Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 1966.

[3] M.A. Djauhari, "*Improved Monitoring of Multivariate Process Variability*", Journal Quality of Technology 37(1), 32-39, 2005.

[4] M.A. Djauhari, "*A Measure of Multivariate Data Concentration*", Journal of Applied Probability & Statistics 2(2), 139-155, 2007.

[5] M.A. Djauhari, *Advanced Monitoring Techniques for Complex Process Variability in Manufacturing Industry*, Universiti Putra Malaysia Press, Serdang Malaysia, 2016.

[6] M.A. Djauhari, M. Mashuri and D.E. Herwindiati, "*Multivariate Process Variability Monitoring*", Communications in Statistics – Theory and Methods 37, 1742-1754, 2008.

[7] M.A. Djauhari, Z. Zolkeply and D.E. Herwindiati, *Reliable Shewhart-type Control Charts for Multivariate Process Variability*, Lambert, Singapore, 2018.

[8] Y. Escoufier, "*Opérateur Associé à un Tableau de Données*", Annales de l'INSEE 22/23, 167-179, 1976.

[9] Y. Escoufier, "*Objectifs et Procédures de l'analyse Conjointe de Plusieurs Tableaux de Données*", Statistique et Analyse des Données 10(1), 1-10, 1985.

[10] F. Glacon, "*Analyse conjointe de plusieurs matrices de données*", PhD Thesis, Université Scientifique et Médicale de Grenoble, 1981.

[11] S. Holmes, "*Multivariate Data Analysis: The French Way*", IMS Collections Probability and Statistics: Essays in Honor of David A. Freedman 2, 219–233, 2008.

[12] J.E. Jackson, *Multivariate quality control: 40 years later, in Statistical Process Control in Manufacturing*, edited by J.B. Keats and D.C. Montgomery, Eds. Marcel Dekker, New York, NY & ASQ Press, Milwaukee, Wisconsin, pp. 123-138, 1991.

[13] C. Lavit, "*Application de la Méthode Statis*", Statistique et Analyse des Données 10(1), 103-116, 1985.
[14] C. Lavit, *Analyse Conjointe de Tableaux Quantitatifs*, Masson, Paris, 1988.
[15] C. Lavit, Y. Escoufier, R, Sabatier and P. Traissac, "*The ACT (STATIS Method)*", Computational Statistics & Data Analysis 18, 97-119, 1994.
[16] H. L'Hermier des Plantes, "*Structuration des tableaux à trois indices de la statistique*". PhD Thesis, Université de Montpellier 2, France, 1976.
[17] H. L'Hermier des Plantes and B. Thiebaut, "*Etude de la Pluviosité au Moyen de la Méthode Statis*", Revue de Statistique Appliquée XXV(2), 57-81, 1977.
[18] R.L. Mason, Y.M. Chou and J.C. Young, "*Monitoring Variation in a Multivariate Process When the Dimension is Large Relative to the Sample Size*", Communications in Statistics-Theory and Methods 38, 939-951, 2009.
[19] D.C. Montgomery, *Introduction to Statistical Quality Control*, Wiley, New York, 2001.
[20] D.C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley and Sons, Inc., New York, 2005.
[21] D.C. Montgomery, *Introduction to Statistical Quality Control*, John Wiley and Sons, Inc., New York, 2009.
[22] R.J. Muirhead, *Aspect of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York, 1982.
[23] T.J. Rato and M.S. Reis, "*Sensitivity Enhancing Transformations for Monitoring the Process Correlation Structure*", Journal of Process Control 24, 905-915, 2014.
[24] N. Razali and B. Wah, "*Power Comparison of Shaphiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests*", Journal of Statistical Modeling and Analytics 2, 21-33, 2011.
[25] M.S. Reis and T.J. Rato, "*Large-Scale Statistical Process Control: Should We Use Partial or Marginal Correlations?*" American Institute of Chemical Engineers (AIChE). Paper presented at the Annual Meeting, November 3–8, 2013.
[26] T.P. Ryan, *Statistical Methods for Quality Improvement*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2011.
[27] J.H. Sullivan, Z.G. Stoumbos, R.L. Mason and J.C. Young, "*Step-Down Analysis for Changes in the Covariance Matrix and Other Parameters*", Journal of Quality Technology 39, 66-84, 2007.
[28] W.H. Woodall and D.C. Montgomery, "*Research Issues and Ideas in Statistical Process Control*", Journal of Quality Technology 31(4), 376-386, 1999.
[29] C-L. Yen and J-J.H. Shiau, "*A Multivariate Control Chart for Detecting Increases in Process Dispersion*", Statistica Sinica 20, 1683-1707, 2010.
[30] C-L.Yen, J-J.H. Shiau and A.B. Yeh, "*Effective Control Chart for Monitoring Multivariate Process Dispersion*", Quality and Reliability Engineering International 28, 409-426, 2012.